RESIDUAL VARIANCE SCALING AND MATRIX APPROXIMATION

Paul Horst
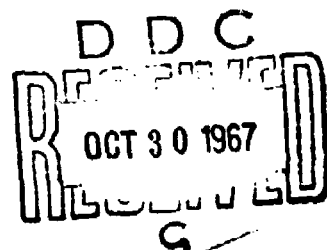
October 1967

University of Washington

Seattle, Washington

D D C

OCT 3 0 1967

CLEARINGHOUSE

RESIDUAL VARIANCE SCALING AND MATRIX APPROXIMATION

Contents

RESIDUAL VARIANCE SCALING AND MATRIX APPROXIMATION

1. Philosophical Orientation

Every human discipline develops terminology and concepts peculiar to its own needs and interests. Terminology developed by a discipline may shape and direct but it can also obscure the basic underlying concepts essential to the development of the discipline. This appears to be true for all human disciplines, whether scientific, political, religious, esthetic, or what not. We must, of course, have verbal, auditory, or other types of symbols to communicate the concepts which are developed within a discipline. Unfortunately, after verbal symbols become established there is often a tendency to confuse them with the fundamental concepts of the discipline. In much of human communication the problem is often one of semantics rather than of agreement as to what are the essential concepts of the discipline.

The confusion between terminology and underlying concepts is not restricted to the nonscientific disciplines. In the sciences as well as the humanities, semantic difficulties are common. Particularly in the sciences where we like to think that our terminology is less ambiguous than in other disciplines, the problems of communication are not confined to the ambiguity of words alone. But even here communication and consequently the development of the science can be either impeded or facilitated by the selection of a particular model or set of underlying philosophical constructs on the basis of which we attempt to regularize observations. These observations may be generated either from events uncontrolled by the observer, such as economics, astronomy, and so on, or by systematically generated experience commonly known as scientific experimentation.

It is important to recognize not only when difficulties of agreement are due to semantic ambiguities but also when they are due to disparities among the underlying philosophical constructs utilized either consciously or unconsciously by the communicators. The problems of semantics and philosophical constructs are perhaps nowhere more pronounced in the scientific disciplines than in the field of psychology. Communication and therefore progress in psychological science can be impeded by preoccupation with both semantics and philosophical models at the expense of more basic issues.

A striking example of how semantic and philosophical ambiguities can cause confusion in a discipline is found in an area of psychology where the techniques of mathematical statistics have been introduced. We refer here to that general field of activity which has come to be known as factor analysis. There is, of course, disagreement as to what specific kinds of activities should be designated as factor analytic. It is perhaps unfortunate that the techniques which have come to be designated as factor analysis have been developed and utilized more extensively in psychology than in other scientific disciplines. One even gets the impression that factor analysis is regarded by some as a branch of psychology. The work of Spearman (1927) in the early part of the century contributed much to this notion that factor analysis is a branch of psychology. It is well known, of course, that his general and specific factor theory of intelligence formed the basis for the numerical and statistical techniques developed to demonstrate his two-factor theory. It is also well known that Thurstone (1947) generalized Spearman's two-factor theory by expanding the general factor into a number of common factors. It is further well known that Hotelling (1933), in an effort to give mathematical elegance to the multidimensional study of

intelligence, developed what has come to be known as principal component analysis. The distinction between Hotelling's principal component analysis and Thurstone's common factor analysis has been the source of much controversy. Perhaps most of this controversy is based on semantic and philosophical preferences rather than on fundamental concepts.

In any case, it has been amply demonstrated over the past several decades that factor analysis is not a branch of psychology, but rather that it is a methodology applicable to all of the sciences. It has not been so clearly demonstrated that factor analysis is a general methodology of which there are many special cases. For example, there are some who would contend that factor analysis is a special case of mathematical statistics. Perhaps the safest way to avoid unproductive semantic and philosophical controversy is to adhere as closely as possible to arithmetical concepts. It is probable that if discussion in any field of human endeavor which purports to be in any sense constructive were confined more closely to arithmetical and numerical considerations, controversy and ambiguity could be greatly reduced. In any case, while the following discussion will be related to what has come to be known as factor analysis methodology in psychology, we shall attempt to adhere as closely as possible to arithmetical concepts and to exclude the more abstract concepts of psychology and mathematical statistics.

In confining our discussion primarily to arithmetical considerations, we exclude also most of mathematics. The reason for this excessive restriction is that even in mathematics, semantic and philosophical red herrings may confuse communication and methodology. It is well known that many different mathematical rationales may lead to the same numerical

results. It is probably in general true that the more ancient a discipline
the more it tends to become encrusted with irrelevant and ritualistic
semantic and philosophical devices. This is true of law, medicine, religion,
philosophy, and mathematics, the last three of which are among the oldest
of formal human disciplines. It is hoped therefore that our presentation
can be maintained almost exclusively on the arithmetical level, and that
even the algebra which it is necessary to employ will be merely shorthand
notation for the arithmetic operations involved. Even though we shall
attempt to restrict the major part of our discussion to numerical concepts,
we shall nevertheless relate the procedures to methods and systems developed
by psychologists and mathematical statisticians. Our own notation and
terminology will follow closely that which we have developed previously
(Horst, 1963, 1965) to circumvent some of the more cumbersome nomenclature
of traditional mathematics.

2. The Arithmetical Model

Suppose we have given an M x n basic data matrix X with $N > n$ and

$$X'1 = 0 \tag{2.1}$$

Consider an approximation matrix U of rank m with $m < n$ such that

$$X - U = \epsilon \tag{2.2}$$

where ϵ is a residual matrix and

$$U' \epsilon = 0 \tag{2.3}$$

Let

$$G = X' X / N \tag{2.4}$$

$$G_u = U' U / N \tag{2.5}$$

$$G_\epsilon = \epsilon' \epsilon / N \tag{2.6}$$

5

From Eqs. 2.2 through 2.6

$$G - G_u = G_\epsilon \qquad (2.7)$$

Let A be an n x m matrix such that

$$G_u = A A' \qquad (2.8)$$

Let

$$D_\epsilon = \text{diag } (G_\epsilon) \qquad (2.9)$$

$$D_A = \text{diag } (G_{AA'}) \qquad (2.10)$$

$$D_G = \text{diag } (G) \qquad (2.11)$$

$$E = D_\epsilon^{-\frac{1}{2}} \epsilon \, D_\epsilon^{-\frac{1}{2}} - I \qquad (2.12)$$

From Eqs. 2.7 through 2.12

$$(D_G - D_A)^{-\frac{1}{2}} (G - A A') (D_G - D_A)^{-\frac{1}{2}} - I = E \qquad (2.13)$$

Note that

$$D_E = 0 \qquad (2.14)$$

Let

$$\alpha = D_\epsilon^{-\frac{1}{2}} A \qquad (2.15)$$

$$C = D_\epsilon^{-\frac{1}{2}} G \, D_\epsilon^{-\frac{1}{2}} \qquad (2.16)$$

From Eqs. 2.13 through 2.

$$C - \alpha \alpha' - I = E \qquad (2.17)$$

Let the basic structure of C be indicated by

$$C = Q \, \delta \, Q' \qquad (2.18)$$

and let

$$Q = \begin{bmatrix} Q_m, & Q_s \end{bmatrix} \qquad (2.19)$$

$$\delta = \begin{bmatrix} \delta_m & 0 \\ 0 & \delta_s \end{bmatrix} \qquad (2.20)$$

where m and s are dimensionality subscripts and

$$m + s = n \tag{2.21}$$

Let

$$\phi = \mathrm{tr}\ E^2 \tag{2.22}$$

We wish now to determine A so that ϕ is a minimum. Equation 2.17 means obviously that the matrix X has been scaled so that the variances of the residual matrix are all unity. The minimization of ϕ in Eq. 2.22 means that the sum of squared correlations for the residual matrix is a minimum.

It is well known that ϕ will be a minimum when

$$\alpha = Q_m\ (\delta_m - I)^{\frac{1}{2}} \tag{2.23}$$

hence for α to be real, the smallest δ_{m_m} in δ_m must be

$$\delta_{m_m} > 1 \tag{2.24}$$

From Eqs. 2.17 through 2.20, and 2.23, it can readily be shown that

$$\phi = \mathrm{tr}\ (\delta_s - I_s)^2 \tag{2.25}$$

Because of Eq. 2.14 we have

$$\mathrm{tr}\ \delta_s = s \tag{2.26}$$

Hence ϕ is simply s times the variance of the s smallest roots or basic diagonal elements of C in Eq. 2.18.

It is of interest to note that because of Eqs. 2.16 through 2.20, and 2.24, we may write

$$D_\epsilon^{-\frac{1}{2}}\ G\ D_\epsilon^{-\frac{1}{2}}\ (D_\epsilon^{-\frac{1}{2}}\ A) = D_\epsilon^{-\frac{1}{2}}\ A\ (I + A'\ D_\epsilon^{-1}\ A) \tag{2.27}$$

or more simply

$$(C - I)\ \alpha = \alpha\ \alpha'\ \alpha \tag{2.28}$$

Equations 2.27 and 2.28 are mathematically equivalent to those given by Lawley (1940), Rao (1955), and others, and usually derived from much more elaborate theoretical constructs. The problem of determining A to minimize ϕ has received much attention by these and other investigators. All methods proposed require iterative procedures beginning with initial estimates of A or functions of its elements. Three major difficulties have been encountered: (1) the determination of suitable initial estimates; (2) excessive computation time, even with electronic computers; (3) so-called improper solutions in which some of the elements of D_e may be negative.

The methods referred to have been insistently designated "factor analysis" to distinguish them from what some writers prefer to call principal component analysis. More specifically, they have been variously called maximum likelihood factor analysis, canonical factor analysis, and maximum determinant factor analysis. We have preferred to circumvent the distinction between factor analysis and principal component analysis and to refer to the algebraic model as a specificity scaling model (Horst, 1965a). It will be noted also that our approach emphasizes the scaling and decomposition of the data matrix rather than of the covariance matrix of the data matrix, although this distinction is not germane to the solution.

3. Computational Rationale

Semantic and philosophical preferences aside, a computational procedure developed by Jöreskog (1966) appears to be the best available to date with reference to the problems of initial estimates, computational speed and accuracy, and proper solutions for residual variances. His development provides significance tests for specific values of m. These tests are based

on the more elaborate philosophical substructure of his model which we do not include in our arithmetical development.

We have previously (Horst, 1965a) presented a computational solution which is a special case of a more general basic structure type solution (Horst, 1965b). The solution cited suffers both from unsatisfactory specifications for the selection of initial values and excessive computation time. It appears, however, to restrict the residual variances to positive values. The method begins with a consideration of the general Gramian matrix G and a factor loading matrix A such that

$$G - A A' = \epsilon \tag{3.1}$$

We determine A in Eq. 3.1 so as to minimize tr ϵ^2. We indicate the basic structure of G by

$$G = Q_m \delta_m Q_m' + Q_s \delta_s Q_s' \tag{3.2}$$

where m and s are dimensionality subscripts which correspond to the first m and last s latent roots and vectors of G.

It is well known that the solution for A of width m which minimizes tr ϵ^2 is

$$A = Q_m \delta_m^{\frac{1}{2}} \tag{3.3}$$

From Eqs. 3.1, 3.2, and 3.3

$$A = G A (A' G A)^{-\frac{1}{2}} h \tag{3.4}$$

where h is an arbitrary square orthonormal matrix. In particular, we may indicate the triangular factoring

$$t t' = A' G A \tag{3.5}$$

Then h in Eq. 3.4 may be such that

$$t'^{-1} = (A' G A)^{-\frac{1}{2}} h \tag{3.6}$$

From Eqs. 3.4 through 3.6

$$A = G A t'^{-1} \tag{3.7}$$

Suppose we choose an arbitrary matrix $_0A$ of width m, subject only to the restriction that $_0A' \, G_m$ is basic. We then write the iteration equations

$$_it \, _it' = \, _iA' \, G \, _iA \tag{3.8}$$

$$_{i+1}A = G \, _iA \, _it'^{-1} \tag{3.9}$$

It has been shown (Horst, 1965b) that $_{i+1}A$ converges to $D_m \, S_m^{\frac{1}{2}}$ and therefore minimizes tr ϵ^2 as i increases without limit.

We have used a modification of this method to solve for A in Eq. 2.13 (Horst, 1935). We let

$$D = (D_G - D_{AA'}) \tag{3.10}$$

$$t \, t' = A' \, D^{-\frac{1}{2}} (D^{-\frac{1}{2}} G D^{-\frac{1}{2}} - I) D^{-\frac{1}{2}} A \tag{3.11}$$

Then

$$D^{-\frac{1}{2}} A = (D^{-\frac{1}{2}} G D^{-\frac{1}{2}} - I) D^{-\frac{1}{2}} A t'^{-1} \tag{3.12}$$

From Eq. 3.11

$$t \, t' = A D^{-1} G D^{-1} A - A' D^{-1} A \tag{3.13}$$

From Eq. 3.12

$$A = (G D^{-1} A - A) t'^{-1} \tag{3.14}$$

We may now let

$$U = D^{-1} A \tag{3.15}$$

$$W = G U - A \tag{3.16}$$

Then from Eqs. 3.13 through 3.16

$$t \, t' = U' \, W \tag{3.17}$$

and

$$\begin{bmatrix} t \\ A \end{bmatrix}^{t'} = \begin{bmatrix} U'W \\ W \end{bmatrix} \tag{3.18}$$

Thus the partial triangular factoring of the supermatrix on the right of Eq. 3.18 yields the factor loading matrix A as the lower submatrix on the left. This leads to the iteration equations

$$_iD = (D_G - D_{iA_iA'})$$ (3.19)

$$_iU = {}_iD^{-1}{}_iA$$ (3.20)

$$_iW = G \, {}_iU - {}_iA$$ (3.21)

$$\begin{bmatrix} {}_it \\ {}_{i+1}A \end{bmatrix} {}_it' = \begin{bmatrix} {}_iU' & {}_iW \\ {}_iW \end{bmatrix}$$ (3.22)

Equations 3.19 through 3.22 constitute in slightly different form and notation those we have previously given for the specificity scaled factor analysis solution (Horst, 1965c). We originally suggested that $_0A$ be taken as the principal axis factor matrix for m factors of the correlation matrix corresponding to G. As is well known, the specificity scaled solution is independent of scale for the original variables and hence the correlation matrix R may, without loss of generality, be taken as G, an arbitrarily scaled covariance matrix. When the principal axis solution is taken for $_0A$, it is obvious from Eq. 2.23 that the number of assumed factors cannot exceed the number of roots of R greater than unity. This restriction is consonant with the recommendations of Kaiser (1960) and others for an upper bound to the number of factors.

Let us now return to Eq. 2.15. From this it can be shown that

$$\begin{bmatrix} D_G + D_{\alpha\alpha'} \end{bmatrix} = \begin{bmatrix} D_G - D_{AA'} \end{bmatrix}^{-1}$$ (3.23)

Let

$$\Delta = D_G + D_{\alpha\alpha'} \tag{3.24}$$

From Eqs. 3.23, 3.24, and 3.12

$$t\,t' = \alpha' \, \Delta^{\frac{1}{2}} \, G \, \Delta^{\frac{1}{2}} \, \alpha - \alpha' \, \alpha \tag{3.25}$$

and

$$\alpha = (\Delta^{\frac{1}{2}} \, (G - I) \, \Delta^{\frac{1}{2}} - I) \, \alpha \, t'^{-1} \tag{3.26}$$

From Eqs. 2.15 and 3.23

$$A = (D_G + D_{\alpha\alpha'})^{-\frac{1}{2}} \, \alpha \tag{3.27}$$

The iterative solution indicated by Eqs. 3.25 and 3.26 shows that because of Eq. 3.24 no iteration can yield a negative Δ, or because of Eq. 3.23, a negative residual variance.

4. Initial Estimates

However, the method previously outlined (Horst, 1965c) does suffer from several weaknesses. First, the principal axis approximation for the $_0A$ matrix as determined from the correlation matrix does not appear to be satisfactory. Second, the iterations converge slowly. Third, there is not adequate assurance that the convergence is to an absolute rather than a local minimum.

To overcome the first objection we take a cue from the image analysis model of Guttman (1953). We consider the residual matrix obtained when each variable is estimated by conventional least square procedures from all the others. The covariance matrix of this residual matrix is well known to be given by

$$G_\varepsilon = D_{G^{-1}}^{-1} \, G^{-1} \, D_{G^{-1}}^{-1} \tag{4.1}$$

and has been called by Guttman (1953) the anti-image matrix.

The covariance matrix of estimated variables is given by

$$G_U = G - 2 \, D_{G^{-1}} + D_{G^{-1}}^{-1} \, G^{-1} \, D_{G^{-1}} \tag{4.2}$$

We seek a scaling of the observed covariance matrix such that the corresponding residual covariances will be unity. This will be the case if we let

$$_0D = D_{G^{-1}} \tag{4.3}$$

For then we have say

$$C = D_{G^{-1}}^{\frac{1}{2}} \, G \, D_{G^{-1}}^{\frac{1}{2}} \tag{4.4}$$

and

$$C^{-1} = D_{G^{-1}}^{\frac{1}{2}} \, G^{-1} \, D_{G^{-1}}^{-\frac{1}{2}} \tag{4.5}$$

It is clear therefore that if a covariance matrix is scaled by the square roots of the diagonals of its inverse, the anti-image matrix of the re-scaled covariance matrix will have unity in the diagonals.

We begin now by rescaling the matrix G as indicated in Eq. 4.5, and let the basic structure of C be

$$C = Q_m \, \mathcal{E}_m \, Q_m{'} + Q_s \, S_s \, Q_s{'} \tag{4.6}$$

We let

$$_C\alpha = Q_m \, (S_m - I)^{\frac{1}{2}} \tag{4.7}$$

$$_0A = (I + D_{_0\alpha_0\alpha'})^{-\frac{1}{2}} {}_0\alpha \tag{4.8}$$

When $_0A$ is used for $i = 0$ in Eqs. 3.19 through 3.22, the value $\emptyset$ for successive iterations drops much more rapidly than when the approximation $_0A$ is based on the largest latent roots and associated vectors of the correlation matrix. For data from Hemmerle (1965), re-analyzed by Jöreskog

(1966), it was also found that with a sufficient number of iterations the value of his criterion and Ψ (our D) values were closely approximated. For this example it appears therefore that the absolute minimum rather than a local minimum was reached. Furthermore, no problems of negative residual variances were encountered although several variables which Jöreskog (1966) found to have Ψ values on the boundary appeared small, as will be subsequently indicated.

5. Iterative Procedures

However, the number of iterations required to achieve Jöreskog's solution for Hemmerle's data was 10,000, and required about 21 minutes on the IBM 7094-MOD 1. It was noted, however, that after about 20 iterations a definite drift appeared to establish itself so that the vectors of differences between successive D vectors decreased slowly. The iteration procedure was therefore modified to take advantage of this regularity as follows:

Let

K_1 be a specified number of iterations

K_2 be a specified number of sets of K_1 iterations

E_1 be a parameter to be empirically determined

E_2 be the minimum value allowed for any element of $_iD$ in Eq. 3.19.

For any iteration i we may calculate the criterion

$$_i\phi = tr\ (_iD^{-\frac{1}{2}}\ (G - _iA\ _iA')\ _iD^{-\frac{1}{2}} - I)^2 \tag{5.1}$$

However, this criterion need be calculated only at prespecified intervals. We proceed as follows:

K_1 iterations are computed of the type 3.19 through 3.22 for the set of K_2 iterations. We calculate

$$U = {}_{K_1}A - (K_1 - 1)^A \tag{5.2}$$

and also ${}_{K_1}\phi$ by means of Eq. 5.1.

We assume now that

$$A = {}_{K_1}A + a \, U \tag{5.3}$$

where $\underline{a}$ is some positive scalar quantity. In particular, we let

$$a = E_1 \, K_2 \tag{5.4}$$

where K_2 is the serial order of the set. From Eq. 5.3 we calculate

$$D = [I - D_{AA'}] \tag{5.5}$$

If no element D_i of Eq. 5.5 is less than E_2, we take A as given in Eq. 5.3 and continue with the next set of iterations. Otherwise we take A as ${}_{K_1}A$, and reduce K_2 to

$$K_2 = K_2 \, / \, n_c \tag{5.6}$$

where n_c is a positive number empirically determined.

We continue in this manner so that for each set of iterations we calculate Eq. 5.2 from the last two iteration cycles of the set and Eq. 5.1 from the last iteration cycle. The value K_2 in Eq. 5.3 increases by 1 for each set of iterations, and the beginning A for the next set of iterations is given by Eq. 5.3 unless a D_i in Eq. 5.5 is less than E_2. In this case, K_2 is first reduced by Eq. 5.6, and the beginning A for the next set of iterations is taken as the last A from the previous set.

Presumably the success of the method depends on the choice of the con-stants K_1, K_2, E_1, and n_c. For seven sets of data of widely differing characteristics, good results were obtained with $K_1 = 10$, $K_2 = 10$, $E_1 = 10$,

and $n_c = 3$. Five of these seven sets have been analyzed by Jöreskog (1966)
but his ϕ value is given for only one of these. Jöreskog gives results
based on a number of different assumed numbers of factors for each set of
data. Since his method is presumably at least as accurate as ours and yields
in addition tests of significance for any assumed number of factors, the
only advantage ours may have is length of time required.

In our method we give only upper and lower bounds for the number of
factors and these are highly tentative. If we let

$$G = D^{\frac{1}{2}}_{R}{}^{-1} \; R \; D^{\frac{1}{2}}_{R}{}^{-1} \tag{5.7}$$

and

$$_m Q \; _m \delta \; _m Q' + \; _s Q \; _s \delta \; _s Q' = G \tag{5.8}$$

then the largest value of m will be such that

$$_m \delta _m > 1 \tag{5.9}$$

and the smallest value such that

$$_m \delta _m + \; _s \delta _s > 2 \tag{5.10}$$

In addition we specify that

$$m \lesssim n/2 \tag{5.11}$$

It should be noted that for the method outlined it is quite possible
for a $_i\phi$ value to be greater than for $_{i-1}\phi$. This can occur after an
acceleration indicated by Eq. 5.3. If the value of $\underline{a}$ is kept sufficiently
small it will not occur, but then the rate of convergence may be unaccept-
ably slow. Our procedure provides for grouping of the successively
calculated ϕ values into sets of K_3 each. In particular we may have $K_3 = 1$.
If the lowest ϕ value in set i is lower than the lowest ϕ in set i + 1, the
routine described is terminated and the A matrix corresponding to the lowest

$\emptyset$ value is taken as the starting point for a final set of iterations without acceleration. This is a sort of polishing operation and it appears that 25 iterations is adequate for the data we have analyzed. If no reversals in $\emptyset$ values are encountered, the routine method continues for some prespecified number of sets, after which the polishing iterations occur.

6. Numerical Results

Results for the seven sets of data we have analyzed are given in Table 1. Each column of the table represents a set of data. The rows are as follows:

Row 1 gives the number of variables in the set.

Row 2 gives the source from which the data were taken.

Row 3 gives the number assigned to the set of data by Jöreskog.

Row 4a gives the smaller number of factors solved for.

Row 4b gives the number of factors solved for by Jöreskog which corresponds most closely to our smaller number.

Row 4c gives the larger number of factors solved for.

Row 4d gives the number of factors solved for by Jöreskog which corresponds most closely to our larger number.

Row 5a gives the $\emptyset/2$ values or half the sum of squared residuals for the smaller number of factors as determined after 400 final polishing iterations and therefore assumed to be very close to the minimum value.

Row 5b gives the $\emptyset/2$ values after 25 polishing iterations for the smaller number of factors with K_1, K_2, and E_1 all equal to 10.

Row 5c is the same as row 5a except that the $\emptyset/2$ values are for the higher number of factors.

Row 5d is the same as row 5b except for the higher number of factors.

Row 6a gives the time in seconds for the accelerated and 25 polishing iterations for the lower number of factors. It does not include the computation time for the initial estimate of $_0A$ nor for input. Perhaps 30 per cent to 50 per cent additional time is required for the initial estimate of $_0A$.

Row 6b gives Jöreskog's time on the CDC 3600 for the nearest corresponding number of factors to those in 6a but does not include input and output time.

Row 6c is the same as row 6a for the higher number of factors.

Row 6d is the same as row 6b for the nearest corresponding number of factors to those in 6c.

It is difficult without actually running Jöreskog's program on the IBM 7094 to compare our time with his. If we take his estimate that the CDC 3600 is about two and a half times as fast as the IBM 7094, it appears that for a maximum of ten sets of accelerating iterations with ten iterations to a set, our method is from three to five times more rapid than Jöreskog's and from 99 to 100 per cent as accurate, depending on the particular set of data and the number of factors solved for. However, we have run our program also on the CDC 3600. Our results indicate that the CDC 3600 is at best only 10 per cent faster than the IBM 7094. If this is correct, then our method is at best only 25 per cent to 100 per cent faster.

Our method does not give the level of significance at which a specified number of factors satisfies the so-called factor analysis model as does Jöreskog's method. If desired, his tests could be added to our program. In this case one would probably begin with our lower bound for the number of factors and proceed first downward and then upward with one less and one additional factor at a time.

It is interesting to note that with Data 3 for 8 factors, the $\phi/2$ value of .05806 is reached after 8 accelerated sets of 10 iterations each and 20 polishing iterations, or a total of 100 iterations, while this criterion is attained only after 6,000 nonaccelerated iterations. Table 2 gives to three decimal places for Data 3 the residual variances scaled back to unit variance for the observed covariance matrix for a number of different cases. The corresponding $\phi/2$ values are given in the last row. Column 1 gives our values for 80 accelerated and 25 polishing iterations. Column 2 gives our values for 6,000 unaccelerated iterations. The $\phi/2$ values for these two columns are the same. Column 3 gives our values for 100 unaccelerated iterations. The $\phi/2$ value is almost 14 per cent greater than for the same number of accelerated iterations of column 1. Column 4 gives our values for 10,000 unaccelerated iterations. Column 5 gives Jöreskog's values. The disparity among all columns except column 3 is doubtless far less than the accuracy of the data would require. Nevertheless the Jöreskog method gives the lowest $\phi/2$ value, .05787. This value was calculated by using the specificity variances which he calculated to three decimal places. Jöreskog's published value for ϕ is .1134 so that his $\phi/2$ is .0567. We cannot account for the discrepancy between this value and our value of .0579 calculated from his unique variances. It is perhaps possible that greater decimal accuracy for the unique variances would have given his ϕ value but only three-place accuracy was available to us.

The ratio of our residual sum of squares to that of Jöreskog is 1.004 and, using 2.5 as the ratio of IBM 7094 MOD 1 to CDC 3600 time, was obtained in less than one-fifth the time. One reason for the rapidity of our method is that an iteration cycle indicated by equations 47 through 51 is many

times faster for a small number of factors than a basic structure solution
for the full covariance matrix. The time of the IBM 7094 MOD 1 for a
15 x 15 matrix with 8 factors is less than .12 seconds for one of our iter-
ations, whereas for the basic structure solution it is about 20 to 30 times
as long. Each Jöreskog iteration requires a basic structure solution.

But even though our results for Data 3 with 8 factors is for all prac-
tical purpose as good as those of Jöreskog and much faster, the superiority
of the method for other numbers of factors for Data 3 and for all of the
remaining sets of data has been demonstrated only for speed and not for
accuracy. Our minimum $\emptyset$'s indicated in Table 1 are probably quite accurate
for the initial $_0A$ matrices on which they are based. Whether, however, these
lead to an absolute as well as a local minimum we have not proved empirically
or theoretically. The application of Jöreskog's method for the other data
would doubtless indicate whether we are close to an absolute minimum for
positive unique variances. This would not, however, prove that our method
for selecting the initial $_0A$ converges in general to an absolute minimum.
That the solution is restricted to positive residual variances we have
already shown.

Even though the iteration cycles for the method we have outlined are
very rapid, columns 1 and 2 of Table 2 indicate that it is primarily the
acceleration feature which is responsible for the speed of the method. This
feature increases the speed of the method by a factor of about 60 for Data 3
with the acceleration parameters used. The question may well be raised
whether other acceleration parameters, or indeed other acceleration strategies,
may increase the rate of convergence appreciably. To date we can only say
that we have experimented with many different combinations of values of

iteration parameters and with other methods of determining the augmentation parameter a throughout the successive iterations. To date we have found no acceleration procedure which is clearly as good or better, from the point of view of speed and accuracy, than the values $K_1 = 10$, $K_2 = 10$, $E_1 = 10$, $n_c = 3$.

It is important in closing to emphasize obvious limitations of the method we have outlined.

(1) We have not proved--and it may well not be true--that in general our method for determining $_0A$ leads to an absolute rather than a local minimum sum of squared residuals.

(2) We have not provided a method for determining the number of factors although Jöreskog's procedure for doing this might be incorporated into ours.

(3) We have by no means exhausted all possibilities for appreciably improving the acceleration strategy.

(4) We do not know how well the acceleration strategy and parameters would work on Gramian matrices in general.

TABLE 1

Summary of Results for Data 1-7

Data	1	2	3	4	5	6	7
1 No. of Var.	5	9	15	17	17	33	9
2 Reference	Thomson(1950)	Horst(1965)	Hemmerle(1965)	Bechtold(1961)	Bechtold(1961)	Lord(1956)	Davis(1944)
3 K.G.J. No.*	**	**	9	3	4	5	9
4 No. of Factors							
4a L_s	2	3	6	6	6	7	1
4b L_{J_s}		3	6	6	6	9	1
4c L_H	3	3	8	7	8	15	4
4d L_{J_H}			8	7	7	11	3
5 Criterion							
5a ϕ_s	.1822	.0370	.2078	.2623	.2517	.8246	.1249
5b ϕ_{J_s}	.1876	.0370	.2104	.2626	.2516	.8246	.1249
5c ϕ_H		.0370	.0579	.1664	.0853	.1674	.1677
5d ϕ_{H_c}	.0750	.0370	.0581	.1666	.0860	.1680	.1677
6 Time in Secs.							
6a T_s	2	2	7	16	12	23	2
6b T_{J_s}			20	33	42	84	3
6c T_H	3	3	16***	17	21	75	4
6d T_{J_H}			43	43	38	111	10

* Number assigned to the data set by Jöreskog (1966)
** Not included in Jöreskog's analyses
*** 21 seconds including calculation of ϕ

TABLE 2

Unique Variances for Data 3

	1 100 Acc.	2 6,000 Un. A.	3 100 Un. A.	4 10,000 Un. A.	5 Jöreskog
1	.263	.262	.238	.262	.263
2	.392	.395	.366	.395	.395
3	.458	.457	.451	.457	.458
4	.090	.086	.238	.083	.080
5	.489	.485	.467	.486	.487
6	.259	.260	.280	.259	.259
7	.014	.010	.173	.006	.005
8	.466	.466	.450	.465	.465
9	.662	.663	.618	.663	.664
10	.010	.042	.143	.037	.029
11	.580	.579	.566	.579	.580
12	.504	.497	.515	.498	.499
13	.738	.736	.721	.737	.738
14	.620	.620	.586	.621	.622
15	.010	.014	.143	.009	.005
$\phi/2$	.05806	.05806	.06586	.05796	.05787

REFERENCES

Bechtold, Harold P. "An Empirical Study of the Factor Analysis Stability Hypothesis," *Psychometrika*, 26, 405-432 (1961).

Davis, F. B. "Fundamental Factors of Comprehension in Reading," *Psychometrika*, 9, 185-197 (1944).

Guttman, Louis. "Image Theory for the Structure of Quantitative Variates," *Psychometrika*, 18, 277-296 (1953).

Hemmerle, W. J. "Obtaining Maximum-Likelihood Estimates of Factor Loadings and Communalities Using an Easily Implemented Iterative Procedure," *Psychometrika*, 30, 291-302 (1965).

Horst, Paul. *Matrix Algebra for Social Scientists*. New York: Holt, Rinehart & Winston, 1963.

_____. *Factor Analysis of Data Matrices*. New York: Holt, Rinehart & Winston, 1965 (Chapter 16, 1965a; Chapter 8, 1965b; Chapter 15, 1965c).

Hotelling, Harold. "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, 24, 417-441, 498-520 (1933).

Jöreskog, K. G. "Some Contributions of Maximum Likelihood Factor Analysis," *Psychometrika*, 32, December (1967).

Kaiser, H. F. "The Application of Electronic Computers in Factor Analysis," *Educational and Psychological Measurement*, 20, 141-151 (1960).

Lawley, D. N. "The Estimation of Factor Loadings by the Method of Maximum Likelihood," *Proceedings of the Royal Society of Edinburgh*, A, 60, 64-82 (1940).

Lord, F. M. "A Study of Speed Factors in Tests and Academic Grades," *Psychometrika*, 21, 31-50 (1956).

Rao, C. R. "Estimation and Tests of Significance in Factor Analysis," *Psychometrika*, 20, 93-111 (1955).

Spearman, Charles. *The Abilities of Man*. New York: Macmillan Co., 1927.

Thomson, Godfrey H. *The Factorial Analysis of Human Ability*. Boston: Houghton Mifflin Co., 1950.

Thurstone, L. L. *Multiple Factor Analysis*. Chicago: University of Chicago Press, 1947.

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
University of Washington, Seattle, Washington	None
	2b. GROUP
	None

3. REPORT TITLE

Residual Variance Scaling and Matrix Approximation

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Does not apply

5. AUTHOR(S) (First name, middle initial, last name)

Horst, Paul

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October, 1967	19	12

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
Office of Naval Research Contract Nonr-477(33)	None
b. PROJECT NO	
None	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None	None
d. None	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Office of Naval Research

13. ABSTRACT

This report treats the canonical factor analysis model as a scale free lower rank approximation model. Computational algorithms are developed which appear to be faster than any currently available. The method is applied to published data and the results are compared with those obtained by other methods.

DD FORM 1473 (PAGE 1)

A-31409

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Canonical Factor Analysis						
Scale Free Factor Analysis						
Maximum Likelihood Factor Analysis						

DD FORM 1 NOV 65 1473 (BACK)